

Journal of Applied Sciences 11 (4): 693-699, 2011
 ISSN 1812-5654 / DOI: 10.3923/jas.2011.693.699
 © 2011 Asian Network for Scientific Information

A Study of Instantaneous Emissions through the Decomposition of Directional Measures for Three-way Contingency Tables with Ordered Categories

¹Ida Camminatiello, ¹Luigi D'Ambra, ²Giovanni Meccariello and ²Livia Della Ragione

¹Department of Matematica e Statistica, University of Naples Federico II, Naples

²Istituto Motori, National Research Council (IM-CNR), Naples

Abstract: The aim of this study is to evaluate the Nitrogen Oxides (NOx) emissions through an accurate analysis of vehicle driving behaviour. For this purpose, a three-way contingency table will be carried out, crossing the NOx emissions, the speed and the acceleration. This contingency table will be analysed by the partition of Marcotorchino index. To complement the survey Ordered Non-Symmetric Correspondence Analysis (ONSCA) will be applied.

Key words: Marcotorchino index, ordered non-symmetric correspondence analysis, kinematic parameters, car emissions, emission models

INTRODUCTION

The Transport activities contribute significantly to air pollutant emissions and the impact on emissions is a key element in the evaluation of any transport policy or plan. They are an important and often dominant source of air pollution with direct and indirect negative impacts on human health. Besides CO₂ emission, other traffic emissions (HC, CO, NOx) have a relevant impact on human health and, for historical towns, on manmade heritage. Its evaluation has therefore gained institutional importance in the International Community, both in USA than in Europe and emerging Asian countries (Andre, 2004; Gholamhassan *et al.*, 2007; Saidur *et al.*, 2008).

Nowadays emission factor models are generally structured in three large groups depending from their general aim. Different space and time scale models are developed for different analysis and traffic impact assessment. National, region and city emission inventory requires overall average emission estimation based on few input parameters as fleet composition and average mission profile, average speed related to average traffic situation. A more detailed analysis requires vehicle driving behaviour represented by the time series of vehicle speed in the link, called hereinafter driving cycle. The assessment of environmental impact of measures regarding road crossing control or development of new infrastructure, requires the analysis of instantaneous emissions of individual vehicles to estimate the effect of planned measure on vehicle speed and acceleration.

COPERT III and IV (Ntziachristos *et al.*, 2006) developed in Europe, as well MOBILE6 (USEPA, 2006;

CARB, 2002) models in the United States, predict emissions having as kinematic input only the average trip speed. These are the so called macro-scale models whose aim is the evaluation of an emission inventory of a large road network or a comparing with a national emission inventory. These models work well for large regional areas and also they could not be utilized for micro scale or instantaneous evaluation. Also they need input kinematic data of variable complexity and are therefore adapted to different usages, for assessing national emissions, as far as for calculating the impact of a local traffic control (Ntziachristos *et al.*, 2006).

In the United States since 1995 to 2005 further developments) CE-CERT at the University of California-Riverside developed the so called Comprehensive Modal Emission Model (CMEM), a modal emission model for light duty and small trucks vehicle (Barth *et al.*, 2002). In this model emissions are calculated as a function of the vehicle operating mode; moreover CMEM is capable to predict second by second emissions and fuel consumption for a wide range of vehicle and technologies category, also for extremely low emitting vehicles (Schulz *et al.*, 2000; Barth *et al.*, 2001).

Recently in Europe the research project ARTEMIS (Assessment and reliability of transport emission models and inventory systems) funded by the European Commission within the 5th Framework Research Programme, aimed at combining the experience from different emission calculation models and experimental research in order to arrive at a harmonised methodology for emission estimates at the national and international level (Hickman and McCrae, 2003). During the activities

of this project, different models are developed and grouped as instantaneous model, traffic situation models and kinematic models (Journard *et al.*, 2006; Hickman and McCrae, 2003).

Kinematic Emission Model (KEM) approach in ARTEMIS was developed by Rapone *et al.* (2008, 2005) and predicts average emissions on a micro-trip (driving pattern) obtaining a detailed description of velocity profile using a multidimensional approach. This kinematic model is defined as a meso-scale emission model and it is to point out that it may require less detailed information with respect to instantaneous model. Also, it is based on the analysis of an existing huge emissions data base (different fleet composition, different driving sub-cycle and laboratories involved in the emissions measurements), coming from ARTEMIS EU project. It has to point out that these physical models are extremely complex and require hours of powerful computer running time, so the models mostly used are based on regression analysis of emission data, collected either in laboratory or on the road. Such models estimate real world emission by average or instantaneous emissions data. KEM is based on the Partial Least Squares Regression. The response considered is the measured emission when the car performs a portion of a trip, called hereinafter Driving Cycle (DC). This unit emission (commonly expressed in g km^{-1}) is a function of total mass and driven distance. Because emissions are generally correlated, a set of four dependent variables (CO , HC , NOx and CO_2) was considered. A consistent set of kinematic parameters is used, to represent the real-world driving behaviour of vehicles in any traffic situation, to develop a KEM. Several explicative variables characterizing the kinematics of Driving Cycles (DC) are used. As a starting selection criterion, instead of considering a broad number of kinematic variables and then selecting the most important variables according to statistical criteria, a minimal set of parameters was assumed on the basis of the physical mechanism of emission production. Two complementary ways of explaining emission variation were considered: the variation in exhaust mass (function of energy spent by the vehicle in a driving cycle) and the frequency of acceleration events at different speeds. Hence, variables were divided into two conceptually meaningful blocks. The first block concerns variables defined from the dynamic vehicle equation, plus idling time to consider emission production during vehicle stand-still. The second block is obtained as transformation of a matrix whose i -th row is the speed/acceleration joint distribution for i -th DC, so each row can be represented by a two-way contingency table. Taking into account the NOx

emissions, a three way contingency table can be obtained crossing the NOx emissions, speed and acceleration.

The innovative aspect of this paper is to study the statistical dependence of the NOx emissions, as dependent variable, on speed and acceleration, explanatory variables, preserving the data structure and respecting the ordinal nature of the variables. For this purpose an integrated approach CATegorical Analysis of Variance (CATANOVA), (Light and Margolin, 1971) and Ordered Non-Symmetric Correspondence Analysis (Beh *et al.*, 2007) will be discussed in the second paragraph. Finally, the proposed approach will be applied to a EURO 3 car fleet in the range of 1200-1400 cc and some results in tabular and graphical forms and index to test whether there exists (or not) an association between the variables, will be presented.

MATERIALS AND METHODS

The analysis proposed in this study is based on an emission database, built within the European project ARTEMIS. The database is relative to a wide collection of automobiles with different technology, powering fuel, homologation and displacement class, tested on a consistent scenario of real driving cycles, representative of vehicle operating conditions in urban, rural and highway roads and different (from congested to rush) traffic situations. A large number of DC's (about 203 cycles divided into 671 sub-cycles) are used for measuring emission factors. Figure 1 shows the speed and $\log(\text{NOx})$ emission profiles of a vehicle, running a specific DC, the ARTEMIS urban DC test. Such kind of time series represents instantaneous profiles object of the study presented in the research.

Several studies demonstrated exhaust emissions of vehicles in real world operation, depend on the instantaneous values of vehicle speed ($v(t)$) and acceleration ($a(t)$). In fact, the engine is designed to supply the vehicle the power required to balance the drag force, which is in turn a function of speed, acceleration and vehicle characteristics (mass, aerodynamics,...). So, instantaneous values of NOx emissions of vehicles are the response, instantaneous values of speed and acceleration are the explicative variables in our analysis. A three-way contingency table for each driving cycle is built, where the value in a cell is the frequency of instantaneous values of vehicle speed ($v(t)$), acceleration ($a(t)$) and $\log \text{NOx}$ emissions detected in the specific class of speed, acceleration and $\log \text{NOx}$ emissions. Table 1 describes the classes. For the definition of the speed, acceleration and $\log \text{NOx}$ categories, the criterion proposed by Andre (2004) is followed.

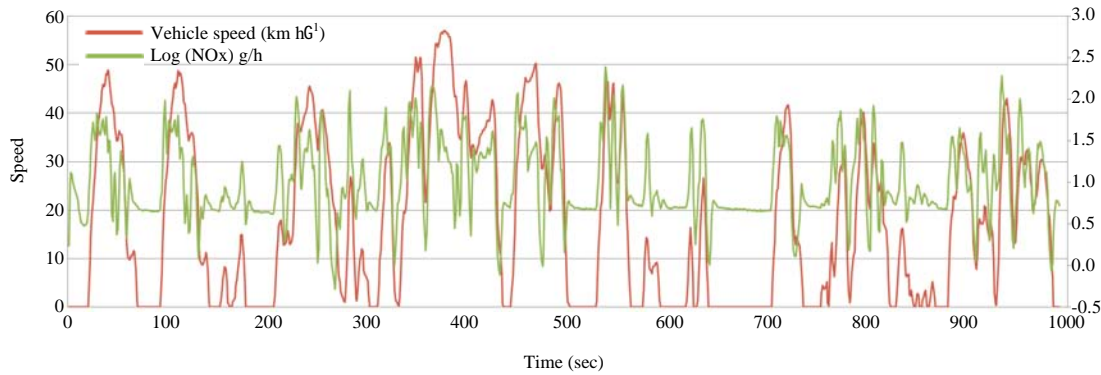


Fig. 1: Speed and log NOx emission profile of test vehicle running Artemis-Urban DC

Table 1: The description of categories

Categories	Speed (km h ⁻¹)	Acceleration (m sec ⁻²)	Log (NOx) (g/s)
0≤speed1≤10	0≤speed1≤10	-∞≤acc1≤-1.4]	nox1<-0.625
10<speed2≤20	10<speed2≤20	-1.4≤acc2≤-0.6	-0.625<nox2<-0.225
20<speed3≤30	20<speed3≤30	-0.6≤acc3≤-0.2	-0.225<nox3<0.475
30<speed4≤40	30<speed4≤40	-0.2≤acc4≤0.2	0.475<nox4<1.175
40<speed5≤50	40<speed5≤50	0.2≤acc5≤0.6	nox5≥1.175
Speed6 > 50	Speed6 > 50	0.6≤acc6≤1.4	
		1.4≤acc7≤+∞	

In order to study the relationship between NOx emissions speed and acceleration, several method can be applied. In the next paragraph we describe some methods for analysing a three-way contingency table, their advantages and disadvantages and propose a combined approach.

A quick look at methods for three-way contingency table:

When the variables are collected in a contingency table, classical statistical tools as correspondence analysis and log linear models are applied.

The aim of correspondence analysis, as well as many multivariate data analytic techniques is to determine scores which describe how different two categories are. To determine the scoring of the rows and columns and the strength of the association between them, the Pearson ratio is partitioned using the method of singular value decomposition.

The log linear analysis focuses on detecting interactions in a multiway contingency table. The basic strategy in loglinear modeling involves fitting models to the observed frequencies in the cross-tabulation of categorical variables. The models can then be represented by a set of expected frequencies that may or may not resemble the observed frequencies. Once the model has been fitted, it is necessary to decide which model provides the best fit. The overall goodness-of-fit of a model is assessed by comparing the expected frequencies to the observed cell frequencies for each model. The

Pearson Chi-squared statistic or the likelihood ratio (L^2) can be used to test a model fit.

In this analysis, both the methodologies can not be used because there is a directional relationship between the variables (one response variable and two predictors). Moreover the chi squared test requires that the expected cell frequencies are not too small (preferably at least five). Instead our contingency table show several cells equal to zero.

The most proper methodology to analyse our data is to partitioning the Marcotorchino index τ_M Eq. 1 for a three-way contingency table with three ordered categorical variables using orthogonal polynomials (Beh *et al.*, 2007). It allows us to study the dependency relationship between the emissions and kinematic variables respecting the asymmetric and ordinal structure of the data and picking up the nonlinear relationship within the data.

Consider a three-way contingency table N that cross-classifies n units according to I row, J column and K tube categories. Denote the (i, j, k) th joint frequency by n_{ijk} with a relative cell frequency of $p_{ijk} = n_{ijk}/n$. Let $p_{i..}$ be the i th row marginal proportion so that

$$\sum_{i=1}^I p_{i..} = 1$$

Similarly let $p_{.j.}$ and $p_{..k}$ be the j th column and k th tube marginal proportions so that

$$\sum_{j=1}^J p_{.j.} = \sum_{k=1}^K p_{..k} = 1$$

Suppose that the relationship between these three variables is such that the J column and K tube categories are predictor variables and are used to predict the outcome of the I row response categories. A measure of predictability can be made by calculating the Marcotochino index, defined as:

$$\tau_M = \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K p_{i,j,k} \left(\frac{p_{ijk}}{p_{i..} p_{.jk}} - p_{i..} \right)^2}{1 - \sum_{i=1}^I p_{i..}^2} \quad (1)$$

It varies from 0 to 1, where, $\tau_M = 0$ indicates there is no dependence, $\tau_M = 1$ indicates there is perfect dependence.

Orthogonal polynomials: The numerator of Marcotorchino can be decomposed using orthogonal polynomials for each of the ordinal-scale variables involved in the partition. The advantage of considering these polynomials is that they reflect sources of category variation. Let us compute the polynomials for the column variables. Let $b_v(j)$ be the v th order polynomial for the j th column category, it can be derived using the general recurrence formula of Emerson (1968):

$$b_v(j) = S_v \left[(s_j(j) - T_v) b_{v-1}(j) - V_v b_{v-2}(j) \right] \quad (2)$$

where:

$$\begin{aligned} T_v &= \sum_{j=1}^J p_{.j} s_j(j) b_{v-1}(j) \\ V_v &= \sum_{j=1}^J p_{.j} s_j(j) b_{v-1}(j) b_{v-2}(j) \\ S_v &= \left\{ \sum_{j=1}^J p_{.j} s_j(j)^2 b_{v-1}^2(j) - T_v^2 - V_v^2 \right\}^{-1/2} \end{aligned}$$

for $v = 0, 1, \dots, J-1$, where, $b_{-1}(j) = 0$, $b_0(j) = 1$ e $s_j(j)$ is the score assigned to j th column category. We consider natural scores such that $S_j(j) = j$ for $j = 1, 2, \dots, J$. These polynomials are subject to the orthonormal constraint:

$$\sum_{j=1}^J p_{.j} b_v(j) b_{v'}(j) = \begin{cases} 1, & v = v' \\ 0, & v \neq v' \end{cases} \quad (3)$$

Similarly, we define $a_u^*(i)$ the u th order polynomial for the i th row category and $c_w(k)$ the w th order polynomial for the k th tube category.

The partition of Marcotorchino's index for a three-way contingency table with three ordinal-scale variables: For three asymmetrically related variables, different decompositions of Marcotorchino numerator are possible depending on the number of ordinal-scale variables. For sake of simplicity, we here illustrate the case where all the three variables are on ordinal-scale. The Marcotorchino numerator, N_m can be partitioned so that:

$$N_m = \sum_{i=1}^{I-1} \sum_{v=1}^{J-1} Z_{iv0}^2 + \sum_{i=1}^{I-1} \sum_{w=1}^{K-1} Z_{i0w}^2 + \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z_{0vw}^2 + \sum_{i=1}^{I-1} \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z_{i vw}^2 \quad (4)$$

where:

$$\begin{aligned} Z_{iv0} &= \sum_{j=1}^J \sum_{k=1}^K a_u^*(i) b_v(j) p_{ijk}, \quad Z_{i0w} = \sum_{j=1}^J \sum_{k=1}^K a_u^*(i) c_w(k) p_{ijk} \\ Z_{0vw} &= \sum_{j=1}^J \sum_{k=1}^K b_v(j) c_w(k) p_{ijk}, \quad Z_{i vw} = \sum_{j=1}^J \sum_{k=1}^K p_{ijk} a_u^*(i) b_v(j) c_w(k) \end{aligned}$$

It is demonstrated (D'Ambra *et al.*, 2005) that the quantity $(n-1) Z_{011}^2$ is a Chi-squared statistic with 1 degree of freedom proposed by Agresti (1990) to measure the linear relationship between the two ordinal variables; the quantity

$$\sum_{j=1}^J Z_{0j1}^2 = Q$$

is the Yates (1948) or Anderson and Landis (1982) statistic. Large value of Q indicates an overall linear trend or mean, compared with overall distribution.

For the sake of simplicity, Eq. 4 can be alternatively expressed as:

$$N_m = \tau_{IJ} + \tau_{IK} + \tau_{JK} + \tau_{IJK} \quad (5)$$

where, τ_{IJ} is the numerator of Goodman and Kruskal (1954) index between the I row response categories and the J column predictor categories, τ_{IK} is the numerator of Goodman-Kruskal index between the I row response categories and the K column predictor categories, τ_{JK} is the Chi-squared statistic between the two predictors, τ_{IJK} is the trivariate association between the response and two predictor variables.

Multiplying each term of N_m by

$$(I-1)(n-1) / 1 - \sum_{i=1}^I p_{i..}^2$$

we obtain C statistics (Light and Margolin, 1971):

$$\begin{aligned} C_M &= \frac{(I-1)(n-1)\tau_{IJ}}{1 - \sum_{i=1}^I p_{i..}^2} + \frac{(I-1)(n-1)\tau_{IK}}{1 - \sum_{i=1}^I p_{i..}^2} + \\ &+ \frac{(I-1)(n-1)\tau_{JK}}{1 - \sum_{i=1}^I p_{i..}^2} + \frac{(I-1)(n-1)\tau_{IJK}}{1 - \sum_{i=1}^I p_{i..}^2} = \\ &= C_{IJ} + C_{IK} + C_{JK} + C_{IJK} \end{aligned} \quad (6)$$

The first term, C_{IJ} , can be compared with the statistic obtained from the chi-squared distribution with $(I-1)(J-1)$ degrees of freedom. Therefore (C_{IJ}) can be used to determine if there is a significant asymmetric association between the row and column categories. The other terms

can be treated in the same manner. Therefore, the Marcotorchino index, τ_M can be used to determine a global association between the three variables.

Ordered three-way non symmetrical correspondence analysis: To graphically describe the dependence structure between the variables, Ordered Non-Symmetric Correspondence Analysis (ONSCA) proposed by Lombardo *et al.* (2007) can be carried out. The ONSCA looks for the orthonormal basis which accounts for the largest part of inertia to visualize the dependence structure between the variables in a lower dimensional space by using orthogonal polynomials. The coordinates for the i th row, j th column and k th tube are:

$$f_{im} = \sum_u a_u^*(i)Z_{um}, g_{jm} = \sum_v b_v(j)Z_{vm}, h_{km} = \sum_w c_w(k)Z_{wm} \quad (7)$$

To complement the correspondence plots, more formal tests of the influence of particular categories may be made by considering the confidence circles for the ONSCA. Lombardo *et al.* (2007) showed that 95% confidence circles for the j explanatory column category represented in a two dimensional non-symmetrical correspondence plot has radii length:

$$r_j^* = \sqrt{\frac{5.99 \left(1 - \sum_{i=1}^I p_i^2 \right)}{p_j (n-1)(I-1)}} \quad (8)$$

The confidence circles, that contain the origin, identify exploratory categories that are not statistically influential in helping to explain the response.

RESULTS AND DISCUSSION

The proposed approach is illustrated by results relative to a EURO 3 car fleet in the range of 1200-1400 cc. Urban driving cycles are considered. We consider a three-way contingency table that cross-classifies the units according to $I = 5$ $J = 7$ and $K = 6$ categories of the variables NOx emissions, acceleration and speed, respectively. We treat the NOx emissions as the response variable.

For our three-way contingency table $N_m = 0,297$, it has an associated C-statistic of $C = 3743.451$ (p-value = 0.000). Therefore we can conclude that the speed and acceleration influence the NOx emissions.

If we take into account the ordinal nature of the three variables, we can partition the Marcotorchino index and test whether there exists (or not) an association between the variables. The results are shown in Table 2. By partitioning C we find that:

Table 2: The Partitions of Marcotorchino numerator and the C-statistics

	IJ	IK	JK	IJK	Marcotorchino num.
τ	0.01	0.12	0.11	0.05	0.29
%Cont.	5.15	40.37	37.63	16.85	100.00
C	192.95	1511.18	1408.61	630.71	3743.45
DoF	24.00	20.00	30.00	120.00	194.00
p-value	0.00	0.00	0.00	0.00	0.00

Table 3: Partition of the C_{IJ}

Source	Statistic-value	DoF	p-value	Percentage
Linear	125.25	6	0.00	65.91
Quadratic	6.90	6	0.09	3.58
Error	60.80	12	0.00	31.51
C statistic	192.95	24	0.00	100.00

Table 4: Partition of the C_{JK}

Source	Statistic-value	DoF	p-value	Percentage
Linear	1296.62	5	0.00	85.80
Quadratic	118.05	5	0.00	7.81
Error	96.51	10	0.00	6.39
C statistic	1511.18	20	0.00	100.00

- Both of predictor variables are statistically significant in influencing the NOx emissions, however the speed is more influential predictor than the acceleration
- There is a statistically significant association between the two predictor variables
- There is an interaction between all three variables

Since, all three variables are statistically related to one another, we can identify whether there is a location, dispersion, or higher order interaction between at least two of the variables. Table 3 shows the partition of the C_{IJ} using orthogonal polynomials. The table highlights that:

- Linear and error components are significant (p-value <0.05)
- The linear (location) component is prevalent

Table 4 shows the decomposition of C_{JK} . In this case all components are significant and the linear is the dominant component.

The element C_{JK} is a symmetric association measure between the two predictor variables so the components are calculated for both rows and the columns. Table 5 and 6 show the decomposition of C_{JK} for the rows and the tubes, respectively. The two tables highlight that all components are significant for both the rows and the columns. There is no a dominant component for the rows, whereas the quadratic is the dominant component for the columns. Moreover, the Agresti statistic is equal 9.45 (p-value = 0.002), so there a significant linear relationship between the speed and acceleration.

When all three variables are considered, we can determine the cause of bivariate associations. Table 7-9 show the decomposition of C_{IJK} for all three variables. We can see that the asymmetric association between the Nox emissions and the acceleration is affected by the

Table 5: Decomposition of C_{JK} for the rows

Source	Statistic-value	DoF	p-value	Percentage
Linear	479.81	6	0.00	34.06
Quadratic	459.00	6	0.00	32.59
Error	469.82	18	0.00	33.35
C statistic	1408.61	30	0.00	100.00

Table 6: Decomposition of C_{JK} for the columns

Source	Statistic-value	DoF	p-value	Percentage
Linear	65.24	5	0.00	4.64
Quadratic	800.71	5	0.00	56.84
Error	542.66	20	0.00	38.52
C statistic	1408.61	30	0.00	100.00

Table 7: Decomposition of C_{JK} : Row-Column

Source	Statistic-value	DoF	p-value	Percentage
Linear	242.38	6	0.00	38.43
Quadratic	149.93	6	0.00	23.77
Error	238.40	12	0.00	37.80
C statistic	630.71	24	0.00	100.00

Table 8: Decomposition of C_{JK} : Row-Tube

Source	Statistic-value	DoF	p-value	Percentage
Linear	97.60	5	0.00	15.47
Quadratic	211.85	5	0.00	33.59
Error	321.26	10	0.00	50.94
C statistic	630.71	20	0.00	100.00

Table 9: Decomposition of C_{JK} : Column-Tube

Source	Statistic-value	DoF	p-value	Percentage
Linear	179.67	6	0.00	28.49
Quadratic	238.52	6	0.00	37.82
Error	212.52	18	0.00	33.70
C statistic	630.71	30	0.00	100.00

significant and dominant linear order differences in the speed (Table 7). Similarly, the quadratic order differences of the NOx emission categories are the most dominant feature that leads to the significant association between the two predictor variables (Table 9) and the non linear order differences of the acceleration categories influence the asymmetric association between the NOx emission levels and the speed (Table 8).

To provide graphical summary of the relationship between the response and explanatory variables one may consider the correspondence plots.

Figure 2 shows the projection of the emission and acceleration categories on subspace spanned by linear and quadratic component. The graph shows a positive relationship between modalities of the two variables. In fact constant or deceleration levels affect low NOx emissions. Similarly, acceleration levels tend to lead to high NOx emissions. In the plot 95% confidence circles have been included. The confidence circles highlight that the levels 4 and 7 of variable acceleration are not statistically significant because their circles involve origin. This suggest to review the distribution in classes.

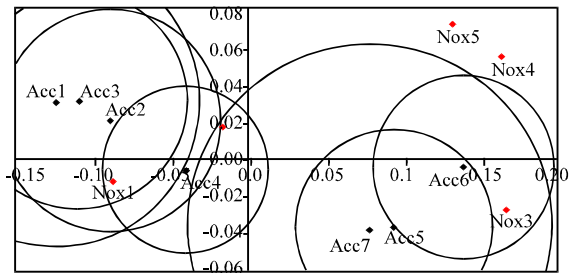


Fig. 2: Non-symmetrical correspondence plot: NOx-acceleration

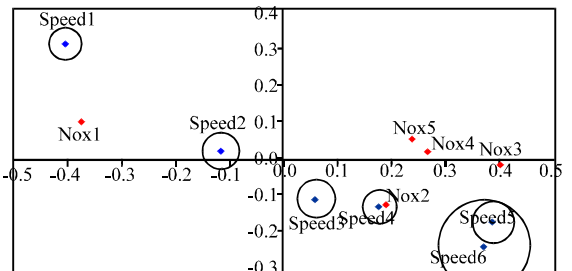


Fig. 3: Non-symmetrical correspondence plot: NOx-Speed

Figure 3 shows the projection of the emission and speed categories on subspace spanned by linear and quadratic component. Also, in this case the plot shows a positive relationship between the categories of two variables. That is low speeds affect low NOx emissions. Similarly, high speeds tend to lead to high NOx emissions. In this case all the categories are statistically significant because no confidence circle involves the origin. Moreover the categories 5 and 6 of variable speed could be unified.

CONCLUSIONS

In this study, we proposed an integrated approach to study the dependence relationship among the NOx emissions and the two independent variables: acceleration and speed. The CATANOVA showed that both acceleration and speed affect in significant manner the NOx emissions. The correspondence plot showed that high speeds and accelerations tend to lead to high NOx emissions. The confidence circles enabled for identifying non significant categories for the response. This suggests that a next survey should review the distribution in classes of the two independent variables.

Our survey can be a valid instrument of decision support indicating the variables which affect the NOx emissions, so we are going to applied the same approach

for the evaluation of CO₂ CO and NO_x emissions in different traffic and road situations.

REFERENCES

- Agresti, A., 1990. Analysis of Categorical Data. Wiley, New-York, USA.
- Anderson, R.G. and R. Landis, 1982. Catanova for multidimensional contingency tables: Ordinal-scale response. *Commun. Statist.*, 11: 257-270.
- Andre, M., 2004. Real-world driving cycles for measuring cars pollutant emissions-Part A: The Artemis European driving cycles. INRETS Report No. LTE 0411, Bron, France.
- Andre, M., 2004. The Artemis European driving cycles for measuring car pollutant emissions. *Sci. Total Environ.*, 334-335: 73-84.
- Barth, M., C. Malcolm, N. Hill and T. Younglove, 2001. Recent validation efforts for a comprehensive modal emissions model. Transportation Research Board 80th Annual Meeting. <http://pubsindex.trb.org/view.aspx?id=694844>.
- Barth, M.J., T. Younglove, T. Malcolm and G. Scora, 2002. Mobile source emissions new generation model: Using a hybrid database prediction technique. Final Report to US. EPA under Award 68-C-01-169. <http://www.epa.gov/oms/models/ngm/cecet.pdf>.
- Beh, E.J., B. Simonetti and L. D'Ambra, 2007. Partitioning a non-symmetric measure of association for three-way contingency tables. *J. Multivariate Anal.*, 98: 1391-1411.
- CARB, 2002. EMFAC2001/EMFAC2002: Calculating emission inventories for vehicles in California: User's guide. California Air Resources Board, Sacramento, CA. <http://www.arb.ca.gov>.
- D'Ambra, L., E.J. Beh and P. Amenta, 2005. Catanova for two-way contingency tables with ordinal variables using orthogonal polynomials. *Comm. Statist. Theory Methods*, 34: 1755-1769.
- Emerson, P.L., 1968. Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, 24: 696-701.
- Gholamhassan, N., G. Barat, Y. Talal and R. Hadi, 2007. Combustion analysis of a CI engine performance using waste cooking biodiesel fuel with an artificial neural network aid. *Am. J. Applied Sci.*, 4: 756-764.
- Goodman, L.A. and W.H. Kruskal, 1954. Measures of association for cross-classifications. *J. Am. Statist. Assoc.*, 49: 732-764.
- Hickman, A.J. and I.S. McCrae, 2003. Revised technical annex, ARTEMIS assessment and reliability of transport emission models and inventory systems. Project Funded by the European Commission within the 5th Framework Research Programme, DG TREN Contract No. 1999-RD.10429. <http://www.trl.co.uk/artemis/introduction.htm>.
- Journard, R., M. Andre, J. Laurikko, T.L. Anh and S. Geivanidis *et al.*, 2006. Accuracy of exhaust emissions measurements on vehicle bench (Artemis deliverable 2). INRETS Report, Bron, France, Report No. LTE 052. http://www.inrets.fr/ur/lte/publications/fichesresultats/ficheartemis/road3/method31/Artemis_deliverable_2_LTE0522.pdf.
- Light, R. and B. Margolin, 1971. An analysis of variance for categorical data. *J. Am. Statist. Assoc.*, 66: 534-544.
- Lombardo, R., E.J. Beh and L. D'Ambra, 2007. Non-symmetric correspondence analysis with ordinal variables using orthogonal polynomials. *Comput. Statist. Data Anal.*, 52: 566-577.
- Ntziachristos, L., C. Kouridis, S. Geivanidis, A. Mamakos and Z. Samaras, 2006. COPERT 4 methodology and software updates. LAT Aristotle University Thessaloniki.
- Rapone, M., M.V. Prati, G. Meccariello, L.D. Ragione and M.A. Costagliola, 2005. A novel statistical model for the evaluation of vehicle emission factors. Application to A Euro III gasoline car fleet. Proceedings of 7th International Conference on Engines for Automobile, ICE005, SAE Paper 2005-24-024. <http://papers.sae.org/2005-24-024>.
- Rapone, M., M.V. Prati, M.A. Costagliola, L.D. Ragione and G. Meccariello, 2008. Emission factors determination of euro III 1,200 to 1,400 cc petrol passenger cars with a PLS multivariate regression model. *Environ. Model. Assess.*, 13: 383-392.
- Saidur, R., M.I. Jahirul, M. Hasanuzzaman and H.H. Masjuki, 2008. Analysis of exhaust emissions of natural gas engine by using response surface methodology. *J. Applied Sci.*, 8: 3328-3339.
- Schulz, D., T. Younglove and M. Barth, 2000. Statistical analysis and model validation of automobile emissions. *J. Transportation Statist.*, 3: 29-38.
- USEPA, 2006. MOBILE model (on road vehicles). Environmental Protection Agency. <http://www.epa.gov/OMS/mobile.htm>
- Yates, F., 1948. The analysis of contingency tables with grouping based on quantitative characters. *Biometrika*, 35: 176-181.